Hexanucleotide Repeats in ALS/FTD Form Length-Dependent RNA Foci, Sequester RNA Binding Proteins, and Are Neurotoxic

Youn-Bok Lee,1 Han-Jou Chen,1 João N. Peres,2 Jorge Gomez-Deza,1 Jan Attig,5,6 Maja Stalekar,2 Claire Troakes,1 Agnes L. Nishimura,1 Emma L. Scotter,1 Caroline Vance,1 Yoshitsugu Adachi,1 Valentina Sardone,1,7 Jack W. Miller,1 Bradley N. Smith,1 Jean-Marc Gallo,1 Jernej Ule,1 Frank Hirth,3 Boris Rogelj,2 Corinne Houart,2 and Christopher E. Shaw1,*

1Department of Clinical Neuroscience, King’s College London, Institute of Psychiatry, London SE5 8AF, UK
2MRC Centre for Developmental Neurobiology, London SE1 1UL, UK
3Department of Biotechnology, Jozef Stefan Institute, Jamova 39, SI-1000 Ljubljana, Slovenia
4Department of Neuroscience, King’s College London, Institute of Psychiatry, London SE5 8AF, UK
5Department of Molecular Neuroscience, UCL Institute of Neurology, Queen Square, London, WC1N 3BG, UK
6MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge CB2 0QH, UK
7Department of Public Health, Neuroscience, Experimental, and Forensic Medicine, University of Pavia, Via Ferrata 9, 27100 Pavia, Italy
*Correspondence: christopher.shaw@kcl.ac.uk
http://dx.doi.org/10.1016/j.celrep.2013.10.049
This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-No Derivative Works License, which permits non-commercial use, distribution, and reproduction in any medium, provided the original author and source are credited.

SUMMARY

The GGGGCC (G4C2) intronic repeat expansion within C9ORF72 is the most common genetic cause of amyotrophic lateral sclerosis (ALS) and frontotemporal dementia (FTD). Intranuclear neuronal RNA foci have been observed in ALS and FTD tissues, suggesting that G4C2 RNA may be toxic. Here, we demonstrate that the expression of 38× and 72× G4C2 repeats form intranuclear RNA foci that initiate apoptotic cell death in neuronal cell lines and zebrafish embryos. The foci colocalize with a subset of RNA binding proteins, including SF2, SC35, and hnRNP-H in transfected cells. Only hnRNP-H binds directly to G4C2 repeats following RNA immunoprecipitation, and only hnRNP-H colocalizes with 70% of G4C2 RNA foci detected in C9ORF72 mutant ALS and FTD brain tissues. We show that expanded G4C2 repeats are potently neurotoxic and bind hnRNP-H and other RNA binding proteins. We propose that RNA toxicity and protein sequestration may disrupt RNA processing and contribute to neurodegeneration.

INTRODUCTION

Amyotrophic lateral sclerosis (ALS) causes progressive muscle weakness and spasticity due to the degeneration of motor neurons. Frontotemporal dementia (FTD) causes changes in personality, language, and behavior due to the degeneration of neurons in frontal and temporal lobes. Both are fatal within 3–5 years of symptom onset. Multiple lines of evidence indicate that these two disorders are phenotypic variants of common pathological processes involving the deposition of TDP-43 (Neumann et al., 2006) or FUS (Vance et al., 2009). A common basis for ALS and FTD came from genetic linkage studies that identified a locus on chromosome 9p21 in familial ALS-FTD (Vance et al., 2006) and genome-wide association studies in sporadic ALS and FTD (Shatunov et al., 2010). The underlying mutation was subsequently shown to be an expansion of the GGGGCC (G4C2) repeat within intron 1 of C9ORF72 (DeJesus-Hernandez et al., 2011; Renton et al., 2011). This mutation accounts for 20%–80% of familial and 5%–15% of sporadic ALS and FTD in North American and European populations (DeJesus-Hernandez et al., 2011; Renton et al., 2011; Smith et al., 2012). The size of the repeat in ALS and FTD cases has been estimated by Southern blotting to range between 700 and 1,600 repeats (DeJesus-Hernandez et al., 2011). The mean number of G4C2 repeats in controls is two; 95% have less than eight repeats (Smith et al., 2012).

The mechanism by which the G4C2 intronic repeats cause neurodegeneration is unknown. Decreased tissue levels of the C9ORF72 transcript implicate a loss of protein function due to haploinsufficiency (DeJesus-Hernandez et al., 2011; Renton et al., 2011). Recent reports describe the aggregation of peptides due to repeat-associated non-ATG (RAN) translation (Mori et al., 2013b) as previously described in SCA8 (Zu et al., 2011). Antibodies against poly (Gly-Ala), (Gly-Pro), and (Gly-Arg) peptides selectively label p62-positive, TDP-43-negative neuronal inclusions that are the pathological hallmark of mutant C9ORF72 cases (Al-Sarraj et al., 2011), but evidence that peptide aggregation initiates neurodegeneration is currently lacking. The identification of intranuclear neuronal RNA foci containing G4C2 repeats in ALS and FTD tissues (DeJesus-Hernandez et al., 2011) is similar to other intronic repeat expansion disorders, including myotonic dystrophy, fragile X tremor ataxia syndrome, and several spinocerebellar ataxias (Todd and Paulson, 2010). RNA foci in myotonic dystrophy sequester and deplete...
muscle-blind-like protein (MBNL1), ultimately causing widespread RNA splicing abnormalities and degeneration of affected tissues (Miller et al., 2000). The fact that overexpression of MBNL1 can rescue the phenotype caused by the CTG repeats implies that loss of MBNL1 is one of the key drivers of cellular degeneration (Kanadia et al., 2003a).

In order to determine whether expanded G4C2 transcripts might be toxic and sequester RNA binding proteins, we generated G4C2 with 8, 38, and 72 repeats and expressed them in cell lines, primary neurons, and zebrafish embryos. Here, we demonstrate that the longer repeat lengths generate RNA foci that are toxic to neurons and bind the RNA binding proteins SF2, SC35, and hnRNP-H. hnRNP-H directly binds to G4C2 RNA and colocalizes closely with RNA foci in transfected cells and the brains of human C9ORF72 ALS and FTD cases. Our findings indicate that G4C2 repeat expansions are potently toxic, sequester RNA proteins, and may initiate neurodegeneration in mutant C9ORF72 ALS and FTD.

RESULTS

Using direct ligation we generated constructs containing 8x, 38x, and 72x G4C2 repeats, which were cloned into an untagged plasmid vector (Figures S1A–S1D). Following transfection into neuronal (SH-SY5Y) cell lines, intranuclear G4C2-positive RNA foci were detected by RNA fluorescence in situ hybridization (FISH) in all cell types expressing 38x and 72x repeats, but not in cells expressing 8x repeats (Figure 1A). The mean number of foci per cell was six for 38x repeats and 12 for 72x repeats (Figure 1E). These were expressed in a range of human nonneuronal cells (CV1, HEK, HeLa) and primary rat cortical neurons (Figures 1F–1I). The RNA foci were resistant to RNase and DNase treatments (Figures S1E and S1F), consistent with the description of G4C2 DNA repeat hairpin loops forming G quadruplexes in vitro (Fratta et al., 2012).

In order to determine whether the length of the G4C2 repeat affected nuclear-cytoplasmic trafficking of the transcript, we cloned the 8x, 38x, and 72x repeat constructs into the 3' position of EGFP expression vectors (Figure S2A). Equivalent transcript expression was detected by RT-PCR and northern blotting for 8x, 38x, and 72x repeat vectors (Figures S2B and S2C). Again, 38x and 72x, but not 8x, repeats or EGFP alone, generated G4C2 RNA foci in SH-SY5Y cell lines (Figures 2A–2D). After 24 hr we observed that a significant proportion of cells with intranuclear RNA foci expressed little or no EGFP (19% ± 1.9% for 38x and 23% ± 1.8% for 72x repeats) (Figure 2E). The absence of EGFP in cells containing foci implies nuclear retention of mRNA containing longer repeats and that foci form shortly after transcription precluding nuclear export. In a time course experiment we observed that the number of SH-SY5Y cells bearing RNA foci decreased dramatically over 72 hr. This was most marked in foci-positive cells that were EGFP negative, such that by 72 hr there was a >90% decrease in the number of cells showing foci for 38x and 72x G4C2 repeats (Figure 2E). Cells with foci that also expressed EGFP (demonstrating nuclear export and cytoplasmic translation) showed a more modest decrease of 30% (EGFP-G4C2 38x) and 21% (EGFP-G4C2 72x), respectively (Figure 2E').

In order to determine whether this decline in numbers was due to toxicity induced by RNA foci, we measured the expression of the apoptosis marker Annexin V by fluorescence-activated cell sorting (FACS). FACS analysis showed that the proportion of EGFP-positive cells expressing Annexin V was 3-fold higher for 38x repeats and 5-fold higher for 72x repeats compared...
to 8× repeats (Figure 2F). To investigate this further, immunocytochemistry (ICC) for activated caspase-3 was performed; it showed that in SH-SY5Y cells, RNA foci dramatically enhanced the activation of caspase-3 (Figure 2G). To quantify the induction of apoptosis by G4C2 foci, we counted caspase-3-activated cells in both foci-negative and foci-positive cell populations. We found no change in the proportion of caspase-3-activated cells above background (3.2% ± 1.2%) in SH-SY5Y cells expressing 8× repeats but a dramatic increase in cells with foci from 38× (37% ± 2.7%) and 72× (34% ± 6.9%) repeats (Figure 2H). Interestingly, HEK293 cells expressing G4C2 38× or 72× showed no evidence of apoptosis despite abundant G4C2 foci (Figure 2H). Accordingly, poly(ADP-ribose) polymerase (PARP) cleavage was detected in 38× or 72× repeat transfected SH-SY5Y cells, but not in HEK293 cells (Figure 2I), suggesting that neuronal cells may be more vulnerable to apoptotic cell death triggered by G4C2 RNA foci.

We next sought to test for length-dependent G4C2 toxicity in vivo by injecting the EGFP constructs with 8×, 38×, and 72× repeats into zebrafish embryos. DNA injection leads to mosaic distribution of the DNA and expression in only a small portion of the embryonic cells. Embryos were injected at the one-cell stage and analyzed at prim-5 (approximately 24 hr post-fertilization). Expression of the EGFP-G4C2 8× repeat did not increase the number of apoptotic cells above EGFP alone (Figures 2J and 2K); however, injection of the 38× or 72× G4C2 repeat constructs (Figures 2L and 2M) resulted in a significant increase in TUNEL-positive cells of 0.8- and 1.4-fold, respectively (graphically represented in Figure 2N). G4C2-positive RNA foci detected by FISH were only observed in 72× embryos, but not in 8× embryos (Figures 2O and 2O′). The majority of 72×-transfected cells that had G4C2-positive RNA foci were also active caspase-3-positive (Figure S2D). TUNEL and active caspase-3-positive cells showed nuclear condensation and fragmentation confirming apoptotic cell death (Figures 2P and 2Q). Thus, evidence from cellular and in vivo assays indicates that longer G4C2 repeats lead to the nuclear retention of transcripts, forming RNA foci that are RNase resistant and potentially toxic.

To investigate whether RNA processing proteins were bound to the G4C2 RNA foci, we probed transfected SH-SY5Y cells with antibodies to 30 different RNA-binding proteins based on their known or predicted ability to bind GC-rich sequences, including heterogeneous nuclear proteins (hnRNP) and serine-arginine (SR) proteins involved in splicing and trafficking (Tables S1 and S2). We found that antibodies to three proteins colocallized with the G4C2 RNA foci: hnRNP-H, serine-arginine-rich splicing factor 1 (SF2), and serine-arginine-rich splicing factor 2 (SC35). hnRNP-H showed the closest association with the foci with almost complete overlap, compared to the distribution of SF2 and SC35 (Figures 3A and 3B). TDP-43 was not mislocalized in any cells with RNA foci. To determine whether these proteins are physically binding to G4C2 RNA, we extracted nuclear lysate from SH-SY5Y cells, incubated it with biotin-labeled G4C2×72 RNA, and performed RNA immunoprecipitation (RIP). Western blots of the eluate were probed for all three proteins, but only hnRNP-H was detected (Figure 3C). To further validate the interaction between G4C2 repeats and hnRNP-H, we performed RIP using rat brain nuclear lysate. hnRNP-H was pulled down from rat brain lysate by G4C2×48 RNA, but not by control RNA of the same length (300 nt), implying that hnRNP-H binds specifically to G4C2 repeats (Figure 3D). Sequestration of hnRNP H by G4C2 RNA foci is expected to reduce the amount of hnRNP-H available and reduce its splicing efficiency on specific RNA targets. hnRNP-H is required for the efficient inclusion of exon 7 into the mature TARBP2 RNA (Xiao et al., 2009). We therefore compared the effect of expanded G4C2 repeats and knockdown of hnRNP-H on TARBP2 exon 7 inclusion. Lentiviral constructs expressing sh-hnRNP-H knocked down hnRNP-H with an efficiency of 95% in SH-SY5Y cells (Figure S3), which dramatically decreased the inclusion of exon 7 into TARBP2 (Figure 3E). A more modest decrease in exon 7 inclusion was also observed in SH-SY5Y cells stably transfected with 72× G4C2

See also Figure S2.
plasmids (Figure 3F). Thus, the presence of G4C2 repeats has a similar effect on the splicing of TARBP2 exon 7 to hnRNPH knockdown and is consistent with the effect of hnRNPH sequestration by G4C2 foci.

Having determined that three RNA binding proteins colocalize with G4C2 RNA foci in cells, we sought a similar association in human tissues. We optimized our FISH protocol (adapted from DeJesus-Hernandez et al. [2011]) and observed abundant RNA foci in the cerebellum and less abundant foci in the frontal cortex and temporal lobes of ALS and FTD patients carrying the C9ORF72 expansion mutation (confirmed by repeat primed PCR and Southern blot, n = 5) that are absent from controls (n = 5) (Figure S4; Table S3). Discrete intranuclear neuronal RNA foci were larger in the cerebellum (~500 nm) compared to the cortex (~200 nm) and were most frequent in neurons adjacent to Purkinje cells (Figures 4A and 4B). By combining FISH for RNA foci and immunofluorescence for the foci-associated RNA binding proteins identified in transfected cells, we demonstrated that SF2 and SC35 rarely colocalized with G4C2 RNA in cerebellum, being present in ~5% of foci, whereas hnRNPH showed striking colocalization overlapping almost with ~70% of foci (Figures 4C–4F).

We have therefore confirmed the presence of RNA foci in ALS and FTD patients with pathological G4C2 intronic expansions in C9ORF72 and demonstrated that they bind RNA-binding proteins, being most closely associated with hnRNPH.

DISCUSSION

The G4C2 expansion mutation in C9ORF72 is the most common gene defect associated with ALS and FTD in the European population. Several mechanistic hypotheses are emerging, including the loss of the C9ORF72 encoded protein, RNA-mediated toxicity (DeJesus-Hernandez et al., 2011; Renton et al., 2011), and toxic dipeptides from repeat-associated non-ATG (RAN) translation (Mori et al., 2013b). Our results provide experimental evidence that RNA containing longer G4C2 repeats causes cellular toxicity in a length-dependent manner in transfected cells and in vivo and may contribute to neurodegeneration in C9ORF72-related ALS and FTD.

Cellular toxicity was associated with the nuclear retention of transcripts containing 38 x and 72 x repeats and the appearance of RNA foci. By placing the repeats 3' to EGFP we were able to monitor nuclear export and show that transfected neuroblastoma cells, but not HEK cells, declined in number over time because of apoptotic cell death. This was most marked in those showing the greatest nuclear retention and foci, indicating cell type and neuron-specific toxicity of longer G4C2 repeats. These data suggest that similar to trinucleotide diseases (Hirth, 2010), cellular toxicity of hexanucleotide repeat expansion directly correlates with G4C2 repeat length, indicating that the longer the repeat, the more toxic its nature.

In order to explore the sequestration hypothesis, we screened antibodies to 30 RNA binding proteins in G4C2 72 x-transfected neuroblastoma cells and demonstrated colocalization of G4C2 RNA foci with SF2, SC35, and, most strikingly, hnRNPH. Over-expression of 72 x G4C2 and knockdown of hnRNPH had a similar effect on the splicing of a target RNA TARBP2, consistent with the sequestration of hnRNPH. Using biotinylated 72 x RNA, we were only able to pull down hnRNPH, indicating that is able to interact with G4C2 RNA. On probing human cerebellar ALS and FTD tissues, we detected a striking colocalization of hnRNPH with 70% of RNA foci. We did not detect changes in the distribution or colocalization with RNA foci in human tissues for pur-a (Xu et al., 2013) or hnRNPH-A3 (Mori et al., 2013a) to suggest sequestration, as had previously been postulated. Our findings suggest that the sequestration of hnRNPH and other RNA binding proteins may play a mechanistic role in neurodegeneration associated with the C9ORF72 mutation.

Our observation that longer stretches of G4C2 RNA form neurotoxic foci and bind specific RNA binding proteins is similar to other intronic microsatellite expansion disorders (Todd and Paulson, 2010). Expanded CUG (DM1) and CCUG (DM2) intronic repeats generate RNA foci in cells, in animal models, and in patients with myotonic dystrophy. The foci bind and deplete muscle-blind-like proteins (MBNL1, 2, and 3), causing widespread RNA splicing abnormalities and degeneration of affected tissues (Miller et al., 2000). The fact that MBNL1 knockout mice develop key aspects of human DM and that overexpression of MBNL1 in Drosophila and mice can rescue the phenotype caused by the CTG repeats implies that loss of MBNL1 is one of the key drivers of cellular degeneration (Kanadía et al., 2003b).
hnRNP-H binds strongly to G-runs in intronic sequences to enhance exon skipping (Xiao et al., 2009) and G-rich RNA quadruplexes (Millevoi et al., 2012). hnRNP-H binding is known to inhibit the nuclear export of RNA containing expanded CUG repeats associated with myotonic dystrophy type 1 (DM1) (Kim et al., 2005). It is a component of the splicing enhancer complex that activates alternative splicing of c-src in neurons (Chou et al., 1999). Thus, the nuclear retention and aggregation of G4C2 RNA may be enhanced by hnRNP-H binding, which could in turn facilitate further G4C2 RNA binding in a positive feedback loop to generate large RNase-resistant foci. This would result in the sequestration of hnRNP-H itself, other RNA binding proteins, and multiple RNA transcripts, leading to significant dysregulation of RNA processing and toxicity. We did not observe TDP-43 binding to G4C2 repeats or mislocalization in our cellular models as has been described in C9ORF72 mutant ALS and FTD cases, and it is not yet clear whether TDP-43 misaccumulation is mechanistic in these cases. The observation that RAN translation generates dipeptides in C9ORF72 mutant cases is of great interest (Mori et al., 2013b). Recent evidence has emerged that RAN polyglycine peptides are present in cellular and Drosophila models of fragile X tremor ataxia syndrome and in patients carrying the CGG expansion and that cell death could be prevented by the suppression of RAN translation (Todd et al., 2013). Toxicity due to RNA foci and RAN translation are not mutually exclusive, and their relative contribution to repeat-associated neurodegenerative disorders requires further investigation.

**EXPERIMENTAL PROCEDURES**

Detailed Experimental Procedures are described in the Supplemental Experimental Procedures.
Patient Samples
All cases were provided by the Medical Research Council (MRC) London Neurodegenerative Diseases Brain Bank (Institute of Psychiatry, King’s College London) and were collected and distributed in accordance with local and national research ethics committee approvals. See Table S3 for case details.

FISH with Immunofluorescence
Cy3-labeled G2/C4 x8 RNA probes were synthesized by Integrated DNA Technologies and used as antisense probes. Cells were fixed in 4% PFA for 15 min and washed with PBS. Then cells were incubated with 70% ethanol overnight (16 hr) at 4°C. The next day, the cells were rehydrated in PBS for 15 min and permeabilized with 0.1% Triton X-100 for 5 min. Following this, the coverslips were incubated in prehybridization solution (40% formamide, 2×SSC) for 15 min. Probes were used at working solutions of 5 ng/μl and diluted in hybridization buffer (40% formamide, 1 mg/ml tRNA, 10% dextran sulfate, 2×SSC). Following overnight hybridization at 48°C, coverslips were washed three times in 2X SSC at room temperature for 15 min each. These were then washed with PBS three times for 15 min each. Coverslips were treated with RNase A/T1 mix (Fermentas, 1:50,000) for 5 min at room temperature and washed with PBS three times for 15 min each. For double immunostaining with FISH, coverslips were blocked in 5% donkey serum for 1 hr, and then cells were incubated with primary antibodies (Table S2), and the appropriate secondary antibodies were used. DAPI (Sigma) was used for counterstaining. Leica confocal SP systems were used for high-resolution imaging.

FISH for Human Tissue
Sections from frontal and temporal cortices and cerebellum were provided as 10% formalin fixed and paraffin-embedded blocks. To perform FISH, paraffin was removed with xylene, and sections were rehydrated in an ethanol series (100%, 95%, 70%) for 3 min per step. These slides were incubated in 0.3% Sudan black for 5 min and washed with water. Next, slides were treated with proteinase-K (20 μg/ml TBS [pH 7.4]) in a 37°C water bath for 20 min. Next, slides were treated with ice-cold 20% acetic acid in 1× TBS for 2 min and then incubated in prehybridization buffer (40% formamide, 2×SSC) for a further 15 min. Probes were used at working solutions of 5 ng/μl and diluted in hybridization buffer (40% formamide, 1 mg/ml tRNA, 10% dextran sulfate, 2×SSC). Following overnight hybridization at 48°C, slides were washed three times in 2X SSC at room temperature for 15 min each. These were then washed with PBS three times for 15 min each.

Statistics
Statistical analysis was carried out using GraphPad Prism 5. For one-way ANOVA, Bonferroni post hoc tests were used to compare the abundance of TUNEL-stained cells in zebrafish. Data are presented as mean ± SEM. Statistical significance was considered at p < 0.0001 (**).


